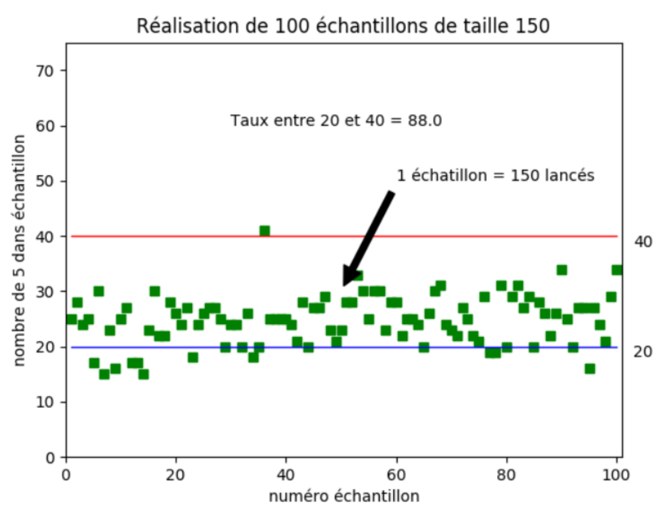
Chapitre 14

Intervalle de Fluctuation
Intervalle de Confiance

14.1 Intervalle de Fluctuation

14.1.1 Exemple par Simulation

- 1 EA consiste à lancer 150 fois 1d6 équilibré
- on note N la VA associée au nombre de fois que le 5 sort
- on cherche $\mathbb{P}(20 \leqslant X \leqslant 40)$
- on simule 100 fois cette EA
- sur 100 expériences simulées, 88 % sont dans l'intervalle [20 ;40], le reste en dehors
- on peut alors dire qu'à 88 %, le nbre d'apparitions de 5 est dans l'intervalle [20 ;40]
- c'est l'***<u>Intervalle de Fluctuation</u>***
  de N au seuil de 88 %



On prendra le temps de bien comprendre cet exemple qui est le coaur de ce chapitre

On pourra faire d'autres simulations avec GeoGebra ou sous Python

14.1.2 IdF et IdF Asymptotique

Définition :

- $X_n \rightsquigarrow \mathscr{B}(n,p)$, $0<\alpha<1$ et $a,b\in\mathbb{R}$
- $[\ a\ ,\ b\ ]$ est un ***Intervalle de Fluctuation*** de X_n au seuil $1-\alpha$ si $\boxed{\mathbb{P}(a\leqslant X\leqslant b)\geqslant 1-\alpha}$

.

ROC - Théorème :

- ***Si :*** $X_n \rightsquigarrow \mathscr{B}(n,p)$
- ***Alors :*** $\forall\, 0<\alpha<1$, $\lim\limits_{n\to+\infty}\mathbb{P}(\dfrac{X_n}{n}\in I_n)=1-\alpha$ ***où :***
  - $\dfrac{X_n}{n}=F_n$ est la fréquence du nombre de succès de la VA binomiale X_n
  - $I_n=[\ p-u_\alpha\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p+u_\alpha\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\ ]$ ***Intervalle de Fluctuation Asymptotique***
  - $u_\alpha>0$ tq $\mathbb{P}(-u_\alpha\leqslant Z\leqslant u_\alpha)\geqslant 1-\alpha$ ***où*** $Z\rightsquigarrow\mathscr{N}(0,1)$
- Animation Approximation faite : Geogebra - IdF et IdFA
- Animation CV (lente) du IdFA vers sa limite p : Geogebra - CV IdFA

.

Preuve :

- $X_n \rightsquigarrow \mathscr{B}(n,p) \quad\Rightarrow\quad \mathbb{E}(X_n)=np$ et $\mathbb{V}(X_n)=np(1-p)$
- ***on pose :*** $Z_n=\dfrac{X_n-np}{\sqrt{np(1-p)}}$ et $Z\rightsquigarrow\mathscr{N}(0,1)$
- d'après Moivre-Laplace, $\lim\limits_{n\to+\infty}\mathbb{P}(-u_\alpha\leqslant Z_n\leqslant u_\alpha)=\mathbb{P}(-u_\alpha\leqslant Z\leqslant u_\alpha)$ <u>avec</u> $u_\alpha>0$
- comme $Z\rightsquigarrow\mathscr{N}(0,1)\Rightarrow\mathbb{P}(-u_\alpha\leqslant Z\leqslant u_\alpha)=1-\alpha$
  $\Rightarrow$ ***d'une part*** , $\boxed{\lim\limits_{n\to+\infty}\mathbb{P}(-u_\alpha\leqslant Z_n\leqslant u_\alpha)=1-\alpha}$
- 1 simple calcul montre que : $-u_\alpha\leqslant\dfrac{X_n-np}{\sqrt{np(1-p)}}\leqslant u_\alpha\Leftrightarrow p-u_\alpha\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\leqslant\dfrac{X_n}{n}\leqslant p+u_\alpha\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}$
  $\Rightarrow$ ***d'autre part*** , $\boxed{-u_\alpha\leqslant Z_n\leqslant u_\alpha\Leftrightarrow\dfrac{X_n}{n}\in I_n}$
- ainsi clairement, $\boxed{\lim\limits_{n\to+\infty}\mathbb{P}(\dfrac{X_n}{n}\in I_n)=1-\alpha}$

.

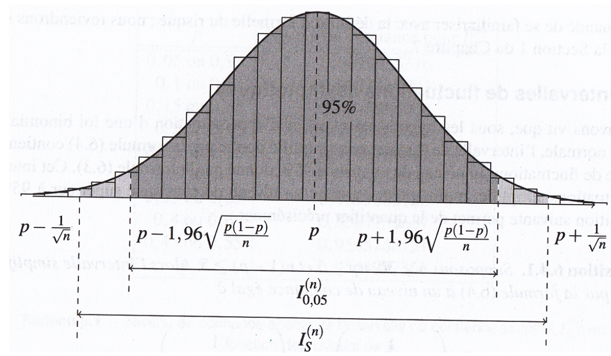
Propriété :

- on retiendra l'***Intervalle de Fluctuation au seuil de 95%***
  - correspondant à $\alpha=0.05$ et $u_\alpha=1.96$: $I_{0.05}^n=[\ p-1.96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p+1.96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\ ]$
  - associé à la VA $F_n=\dfrac{X_n}{n}$
- la notation précise $I_{0.05}^n(\dfrac{X_n}{n})$ est lourde : on note donc en général $I_{0.05}$ mais attention ...

Exemple, remarque :

- on reprend l'exemple sur les 120 lancers de 1d6 avec la VA N (nbre de 6)
- l'intervalle de fluctuation asymptotique au seuil de 95 % (dans les conditions d'approximation) est :
  - $p - 1.96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} = \dfrac{1}{6} - 1.96\dfrac{\sqrt{\frac{1}{6}(1-\frac{1}{6})}}{\sqrt{120}} \approx 0.100$
  - $p + 1.96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} = \dfrac{1}{6} + 1.96\dfrac{\sqrt{\frac{1}{6}(1-\frac{1}{6})}}{\sqrt{120}} \approx 0.233$
  - $I_{0.05}^{120}(\dfrac{N}{120}) = [\ 0.100\,;\, 0.233\ ] \Rightarrow \boxed{I_{0.05}^{120}(N) = [\ 12\,;\, 28\ ]}$ (car $0.100 \times 120 = 12$ et $0.233 \times 120 = 28$)

- $I_{0.05}$ pourrait-être simplifié par l'IdF de 2$^{\text{nd}}$:
  $I_S = [\ p - \dfrac{1}{\sqrt{n}}\,;\, p - \dfrac{1}{\sqrt{n}}\ ]$
- trouver I_S ; que constatez-vous ?
- mq, d'1 façon générale, $I_S \subset I_{0.05}$
- $\boxed{I_{0.05} \text{ de T}^{\text{ale}} \text{ est donc plus précis que } I_S \text{ de 2}^{\text{nd}}}$
  (voir graphe)



- Animation comparant les ***intervalles de fluctuation*** : Geogebra - Idf 2$^{\text{nd}}$, 1$^{\text{ère}}$ et T$^{\text{ale}}$

14.1.3 Prise de Décision

Propriété :

- f_{obs} la fréquence observée d'1 caractère sur un échantillon de taille n issu d'1 population
- on suppose les conditions de l'approximation normale de la loi binomiale sont remplies
  ($n > 30$, $np > 5$ et $n(1-p) > 5$)
- ***Test d'hypothèse :*** on fait une $\boxed{\text{hypothèse sur la valeur de la proportion } p}$ du caractère étudié dans la population toute entière
- I_n l'intervalle de fluctuation asymptotique au seuil de 95 %
- ***Si*** $f_{obs} \in I_n$ ***Alors*** on ne peut rejeter l'hypothèse faite sur p
- ***Si*** $f_{obs} \notin I_n$ ***Alors*** on rejette l'hypothèse faite sur p

.

Exemple, remarque :

- pour créer ses propres colliers, on achète 1 kit de perles :
  - ce kit contient 5 couleurs (marron, jaune, rouge, vert et bleu)
  - les marrons et jaunes sont annoncées comme représentant chacune 20 % de l'ensemble des perles tandis que les rouges sont annoncées à 10 %
- ***on veut vérifier cette information :*** pour cela, on observe un échantillon aléatoire de perles puis on construit un intervalle de fluctuation asymptotique au seuil de 95 % pour la proportion de marron
- on constitue donc 1 échantillon, considéré aléatoire, de 690 perles :
  - on dénombre 140 perles marrons
  - la prise de décision est la suivante : si la proportion de perles marrons dans l'échantillon n'appartient pas à intervalle de fluctuation, on rejette l'hypothèse selon laquelle les perles marrons représentent 20 % des perles
- déterminer l'IdF asymptotique I au seuil de 95 proportion de perles marron. (rép : I = [0, 17 ; 0, 23])
- calculer la proportion de perles marrons dans l'échantillon. Conclusion ?
  (rép : marron ok)
- dans le même échantillon, il y avait 152 perles jaunes et 125 perles rouges. Conclusion ?
  (rép : jaune ok - rouge rejet)

14.2 Estimation

14.2.1 De quoi s'agit-il ?

- pour des raisons de coût et de faisabilité, on ne peut étudier 1 certain caractère sur l'ensemble d'1 population : ***la proportion*** p ***est inconnue***
- ***on cherche alors à estimer*** p à partir d'1 échantillon de taille n
- on calcule la fréquence f_{obs} des individus de cet échantillon ayant ce caractère
- ***on estime la proportion*** p par 1 ***intervalle de confiance*** déterminé à partir de la fréquence f_{obs} et de la taille n de l'échantillon

14.2.2 Intervalle de Confiance

ROC - Propriété - Définition :

- $X \rightsquigarrow \mathscr{B}(n,p)$ où p est inconnu
- F_n la VA qui à chaque échantillon de taille n associe la fréquence du caractère de cet échantillon
- ***Si*** les conditions de l'approximation normale sont remplies :
  $n > 30$, $np > 5$ et $n(1-p) > 5$
- ***Alors*** p vérifie $\mathbb{P}(F_n - \frac{1}{\sqrt{n}} \leqslant p \leqslant F_n + \frac{1}{\sqrt{n}}) \geqslant 0.95$
- en considérant f_{obs}, la fréquence observée pour un échantillon, on pose alors la définition suivante :
  $[\ f_{obs} - \frac{1}{\sqrt{n}}\ ;\ f_{obs} + \frac{1}{\sqrt{n}})\ ]$ est un ***Intervalle de Confiance*** de p au seuil $1-\alpha$
- Animation détermination de p par IdC : Geogebra - recherche d'1 proportion par IdC

Preuve :

- les conditions d'approximation étant vérifiées, on sait que (en considérant l'IdF simplifié) :
  $\mathbb{P}(p - \frac{1}{\sqrt{n}} \leqslant F_n \leqslant p + \frac{1}{\sqrt{n}}) \geqslant 0.95$
- comme $p - \frac{1}{\sqrt{n}} \leqslant F_n \leqslant p - \frac{1}{\sqrt{n}} \Leftrightarrow ... \Leftrightarrow F_n - \frac{1}{\sqrt{n}} \leqslant p \leqslant F_n - \frac{1}{\sqrt{n}}$
- on déduit que : $\mathbb{P}(F_n - \frac{1}{\sqrt{n}} \leqslant p \leqslant F_n + \frac{1}{\sqrt{n}}) \geqslant 0.95$

Exemple : un sondage pour l'élection présidentielle du 21 avril 2002

- voici les résultats d'un sondage IPSOS réalisé avant l'élection présidentielle de 2002 pour Le Figaro et Europe 1, les 17 et 18 avril 2002 auprès de 989 personnes, constituant un échantillon national représentatif de la population française âgée de 18 ans et plus et inscrite sur les listes électorales
- on suppose cet échantillon constitué de manière aléatoire (même si en pratique cela n'est pas le cas)
- les intentions de vote au premier tour pour les principaux candidats sont les suivantes :
  - Jacques Chirac : 20 %
  - Lionel Jospin : 18 %
  - Jean-Marie Le Pen : 14 %
- les médias se préparent pour un second tour entre Jacques Chirac et Lionel Jospin
- déterminer pour chaque candidat, l'intervalle de confiance au niveau de confiance de 0.95 de la proportion inconnue d'électeurs ayant l'intention de voter pour lui.
- le 21 avril, les résultats du premier tour des élections sont les suivantes :
  - Jacques Chirac : 19,88 %
  - Lionel Jospin : 16,18 %
  - Jean-Marie Le Pen : 16,86 %
- les pourcentages de voix recueillies par chaque candidat sont-ils bien dans les intervalles de confiance précédents ?
- pouvait-on, au vu de ce sondage, écarter avec un niveau de confiance de 0.95, l'un de ces trois candidats second tour ?

14.2.3 Synthèse : Quelques vidéos simplifiées mais utiles

- ***Exemple 1 :*** ex vidéo corrigé : Utube - déterminer un IdF
- ***Exemple 2 :*** ex vidéo corrigé : Utube - prendre une décision à l'aide d'un IdF
- ***Exemple 3 :*** ex vidéo corrigé : Utube - estimer une proportion par IdC
- ***Exemple 4 :*** ex vidéo corrigé : Utube - déterminer la taille d'1 échantillon
- ***Exemple 5 :*** ex vidéo corrigé : Utube - distinction IdF et IdC

14.3 Sujet de Bac

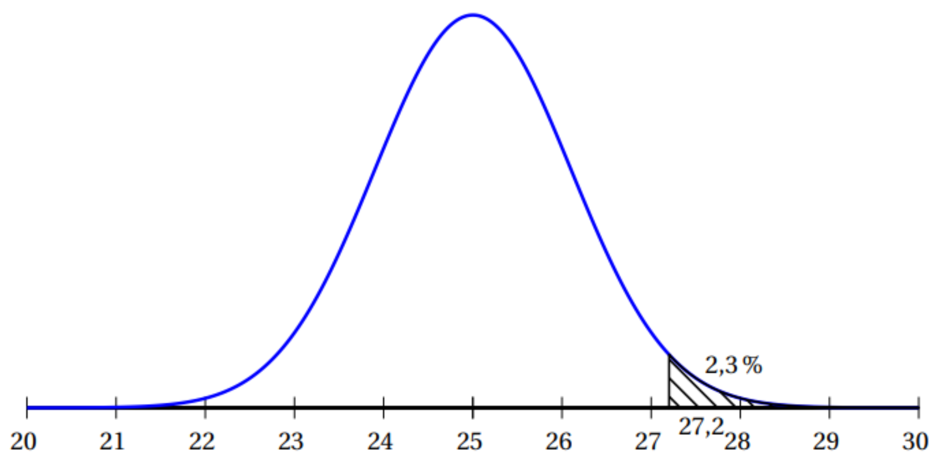
Ex1 : Antilles - Guyane 2017

Dans une usine automobile, certaines pièces métalliques sont recouvertes d'une fine couche de nickel qui les protège contre la corrosion et l'usure. Le procédé utilisé est un nickelage par électrolyse.

On admet que la variable aléatoire X, qui à chaque pièce traitée associe l'épaisseur de nickel déposé, suit la loi normale d'espérance $\mu_1 = 25$ micromètres (μm) et d'écart type σ_1.

Une pièce est conforme si l'épaisseur de nickel déposé est comprise entre 22,8 µm et 27,2 µm.

La fonction de densité de probabilité de X est représentée ci-dessous. On a pu déterminer que $P(X > 27,2) = 0,023$.



1. **a.** Déterminer la probabilité qu'une pièce soit conforme.

   b. Justifier que 1,1 est une valeur approchée de σ_1 à 10^{-1} près.

   c. Sachant qu'une pièce est conforme, calculer la probabilité que l'épaisseur de nickel déposé sur celle-ci soit inférieure à 24 μm. Arrondir à 10^{-3}.

2. Une équipe d'ingénieurs propose un autre procédé de nickelage, obtenu par réaction chimique sans aucune source de courant. L'équipe affirme que ce nouveau procédé permet théoriquement d'obtenir 98 % de pièces conformes.

   La variable aléatoire Y qui, à chaque pièce traitée avec ce nouveau procédé, associe l'épaisseur de nickel déposé suit la loi normale d'espérance $\mu_2 = 25$ μm et d'écart-type σ_2.

   a. En admettant l'affirmation ci-dessus, comparer σ_1 et σ_2.

   b. Un contrôle qualité évalue le nouveau procédé ; il révèle que sur 500 pièces testées, 15 ne sont pas conformes.

   Au seuil de 95 %, peut-on rejeter l'affirmation de l'équipe d'ingénieurs ?

Ex2 : Centres Étrangers 2017

Cet exercice est un questionnaire à choix multiples (Q.C.M.).
Pour chacune des questions, une seule des quatre réponses est exacte. Le candidat indiquera sur sa copie le numéro de la question et la lettre correspondant à la réponse exacte.
Aucune justification n'est demandée. Une réponse exacte rapporte un point, une réponse fausse ou une absence de réponse ne rapportent aucun point.

On étudie la production d'une usine qui fabrique des bonbons, conditionnés en sachets.

On choisit un sachet au hasard dans la production journalière. La masse de ce sachet, exprimée en gramme, est modélisée par une variable aléatoire X qui suit une loi normale d'espérance $\mu = 175$. De plus, une observation statistique a montré que 2 % des sachets ont une masse inférieure ou égale à 170 g, ce qui se traduit dans le modèle considéré par : $P(X \leqslant 170) = 0,02$.

Question 1 : Quelle est la probabilité, arrondie au centième, de l'évènement « la masse du sachet est comprise entre 170 et 180 grammes » ?

Réponse a : 0,04
Réponse b : 0,96
Réponse c : 0,98
Réponse d : On ne peut pas répondre car il manque des données.

Les différents bonbons présents dans les sachets sont tous enrobés d'une couche de cire comestible.
Ce procédé, qui déforme certains bonbons, est effectué par deux machines A et B.
Lorsqu'il est produit par la machine A, la probabilité qu'un bonbon prélevé aléatoirement soit déformé est égale à $0,05$.

Question 2 : Sur un échantillon aléatoire de 50 bonbons issus de la machine A, quelle est la probabilité, arrondie au centième, qu'au moins 2 bonbons soient déformés ?

Réponse a : 0,72
Réponse b : 0,28
Réponse c : 0,54
Réponse d : On ne peut pas répondre car il manque des données

La machine A produit un tiers des bonbons de l'usine. Le reste de la production est assuré par la machine B.
Lorsqu'il est produit par la machine B, la probabilité qu'un bonbon prélevé aléatoirement soit déformé est égale à $0,02$.
Dans un test de contrôle, on prélève au hasard un bonbon dans l'ensemble de la production. Celui-ci est déformé.

Question 3 : Quelle est la probabilité, arrondie au centième, qu'il soit produit par la machine B ?

Réponse a : 0,02
Réponse b : 0,67
Réponse c : 0,44
Réponse d : 0,01

La durée de vie de fonctionnement, exprimée en jour, d'une machine servant à l'enrobage, est modélisée par une variable aléatoire Y qui suit la loi exponentielle dont l'espérance est égale à 500 jours.

Question 4 : Quelle est la probabilité, arrondie au centième, que la durée de fonctionnement de la machine soit inférieure ou égale à 300 jours ?

Réponse a : 0,45
Réponse b : 1
Réponse c : 0,55
Réponse d : On ne peut pas répondre car il manque des données

L'entreprise souhaite estimer la proportion de personnes de plus de 20 ans parmi ses clients, au niveau de confiance de 95 %, avec un intervalle d'amplitude inférieure à $0,05$. Elle interroge pour cela un échantillon aléatoire de clients.

Question 5 : Quel est le nombre minimal de clients à interroger ?

Réponse a : 40
Réponse b : 400
Réponse c : 1600
Réponse d : 20

